A Comparison of Machine Learning Methods to Predict Hospital Readmission of Diabetic Patient

Le Dinh Phu Cuong*

College of Computer Science and Electronic Engineering
Hunan University
Changsha 410082
China
cuongldp@yersin.edu.vn

Dong Wang

College of Computer Science and Electronic Engineering
Hunan University
Changsha 410082
China
wangd@hnu.edu.cn

Corresponding Author email: cuongldp@yersin.edu.vn

ABSTRACT

Diabetes is a chronic disease whereby blood glucose is not metabolized in the body. Electronic health records (EHRs) (Yadav, P. et al., 2018). for each individual or a population have become important to standing developing trends of diseases. Machine learning helps provide accurate predictions higher than actual assessments. The main problem that we are trying to apply machine learning model and using EHRs that combines the strength of a machine learning model with various features and hyperparameter optimization or tuning. The hyperparameter optimization (Feurer, M., 2019) uses the random search optimization which minimizes a predefined loss function on given independent data. The evaluation on the method comparisons indicated that machine learning models has increased the ratio of metrics compared to previous models (Accuracy, Recall, F1 and AUC score) on the same public dataset that is reprocessed.

Keywords: *Amazon Sage Maker (AWS), Machine/Deep Learning, Diabetes*

RESEARCH HIGHLIGHTS

This article presents machine learning as an effective approach to predict hospitalization in diabetics (Muhammad, L. J. et al., 2020)

Besides, we use a XGBoost model (Chen, T., 2015) which also is implemented in Amazon's Sage Maker machine learning platform is slightly higher in the Train Accuracy and the Test Accuracy because the Tuning Hyperparameter focuses intensively on minimizing the Validation Error.

A comparison of Machine/Deep Learning Methods: Logistic Regression, ANN, Naïve Bayesian Classifier, SVM, GBoost (XGBoost)

_____

Research Objectives

Diabetes has a significant incidence of illness, health care effects and lead to a high mortality rate due to its complications in both developed and developing countries. Diabetes treatments are not completely adequate and costly, so we recommend that prevention is an important factor in reducing the burden of diabetes and its complications. We used the T2D dataset by numerous US hospitals.

Machine learning (ML) is one of the important parts of artificial intelligence, provides the methods and techniques to learn from experience. Supervised learning technique involves learning from training data. Unsupervised learning, on the other hand, the system may try to find hidden data patterns, links between features or variables. The reinforcement learning system tries to learn through interactions with dynamic environments in which computer programs provide access to dynamic environments to realize a specific goal.

The goal of this paper applies ML technology for predicting the likelihood of readmission of patients to hospitals. It includes multiple analysis techniques to predict and document the lack of adequate comparative studies to support the selection of a suitable technique.

Methodology

This method used the dataset and five basic ML techniques. That is a comparison of Machine/Deep Learning Methods: Logistic Regression, ANN, Naïve Bayesian Classifier, SVM, GBoost (XGBoost).
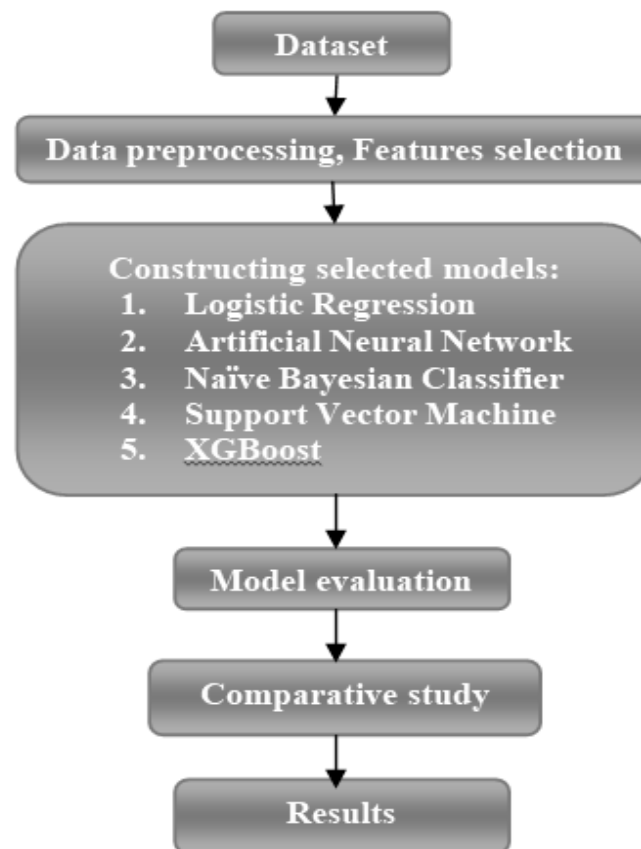

Fig. 1. Methodology Diagram

Results

The readmissions increased costs of health care and negatively affect the reputation of the hospital. Therefore, it is important to predict the number of hospitalizations for diabetics. This article presents machine learning as an effective approach to predict hospitalization in diabetics. Besides, XGBoost model is implemented in Amazon's SageMaker machine learning platform (Joshi, A. V., 2020) is slightly higher in the Train Accuracy and the Test Accuracy because the Tuning Hyperparameter focuses intensively on minimizing the Validation Error. It is better for avoiding 'overfitting' and applying in real-life problems. SageMaker configures tuning job quickly thanks to its integrated Tuning Hyperparameter function. Besides, it allows developers use the Python SDK to impove more in the future. Since then, we will develop cloud computing infrastructure for implementing microservice architecture in developing countries (Cuong, L. D. P. et al., 2020). Furthermore, we also want to mention people to raise awareness of preventing diabetes to avoid serious complications and reduce spending (Chinh, N. V. et al., 2020).
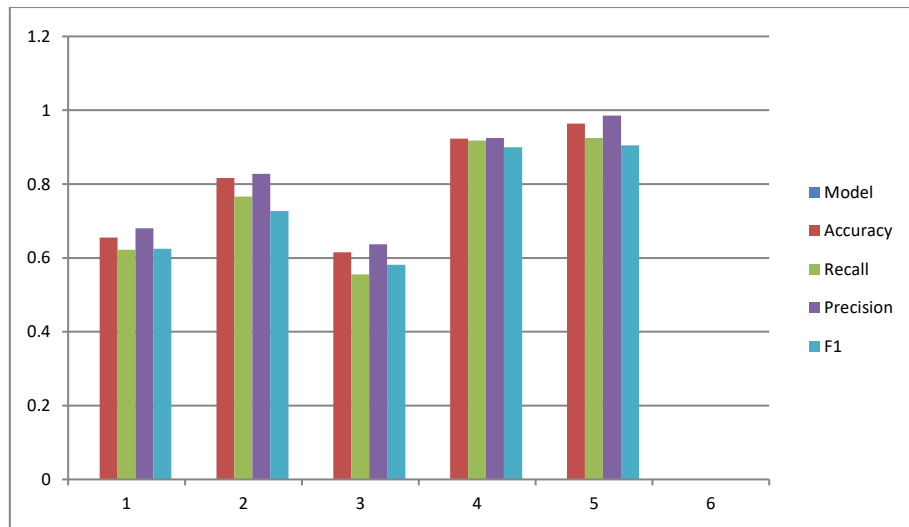
Fig. 2. Results of model evaluation

References

Chen, T., He, T., Benesty, M., Khotilovich, V., & Tang, Y. (2015). Xgboost: extreme gradient boosting. *R package version 0.4-2*, 1-4.

Chinh, N. V., Kiet, P. H. T., Diep, P. B., Yen, N. T. B., Wang, D., Cuong, L. D. P., & Tu, N. H. (2020). Prevention of complicationsand several other relevant risk factors of Type 2 Diabetes (T2D) at a hospital in Vietnam. *European Journal of Molecular & Clinical Medicine*, 7(10), 266-279.

Cuong, L. D. P., Wang, D., et al. (2020). Advanced cloud computing services: the evaluation of a development roadmap for emerging fields in vietnam. J*ournal of critical reviews*, 7(17), 3085-3105.

Feurer, M., & Hutter, F. (2019). Hyperparameter optimization. *In Automated Machine Learning* (pp. 3-33). Springer, Cham.

Joshi, A. V. (2020). Amazon's Machine Learning Toolkit: Sagemaker. *In Machine Learning and Artificial Intelligence* (pp. 233-243). Springer, Cham.

Muhammad, L. J., Algehyne, E. A., & Usman, S. S. (2020). Predictive supervised machine learning models for diabetes mellitus. *SN Computer Science*, 1(5), 1-10.

Yadav, P., Steinbach, M., Kumar, V., & Simon, G. (2018). Mining Electronic Health Records (EHRs) A Survey. *ACM Computing Surveys (CSUR)*, 50(6), 1-40.

Author's Biography

Le Dinh Phu Cuong *is currently a Ph.D candidate in Computer Science from College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China. He is also a lecturer in Department of Information Technology, Yersin University, Vietnam. He received his Master*

degree in Information Technology, Paris VI University. His research interests include Cloud Computing, Wireless Network, Data analytics and Artificial Intelligence. Contact him at: cuongldp@yersin.edu.vn and ledinhphucuong.dalat@gmail.com.

Dong Wang *received the B.S. and Ph.D. degrees in computer science from Hunan University, in 1986 and 2006, respectively. From 2004 to 2005, he was a Visiting Scholar with the University of Technology Sydney, Australia. Since 1986, he has been with Hunan University, China, where he is currently a Professor. His main research interests include network test and performance evaluation, wireless communications, and mobile computing. Contact him at: wangd@hnu.edu.*